From novice to ninja

Tools and techniques for a productive career in scientific computing

Ignacio Romero
ignacio.romero@imdea.org

Technical University of Madrid, Spain
IMDEA Materials Institute, Madrid, Spain

January 30th, 2023

Outline

1. Motivation

Goals

- Introduce general principles that should guide your efforts when choosing what software to employ.
- Describe *tools* that are useful to develop a career in applied math, computational mechanics, computational materials science, computer science, etc.
- Do it from the simplest (a must!) to the more advanced ones (beware!).
- Give examples to illustrate that the effort pays off.
- Provide links for exploration (see pdf file in group web).
- Open the floor to discussion.

Disclaimers

- Everything is arguable: this is just an over-simplified presentation.
- Everything is subject to opinion.
- The goal is more important than the tool.
- I am biased in favor of free and open-source software.
- I might make use of humour and exaggeration to stress ideas.
- I do not claim to be a "black belt". I'm on the path of the ninja ...
- You do not need these tools to be successful in research ... but might help.

General principles

- Every day there are new, flashier, more powerful computer programs.
- Artificial intelligence is enabling incredible applications.
- User interfaces, web based, social media-based frameworks are alluring.

General principles

- Every day there are new, flashier, more powerful computer programs.
- Artificial intelligence is enabling incredible applications.
- User interfaces, web based, social media-based frameworks are alluring.
- But every new tool demands some effort
- And you do not want to learn yet another tool that might not be available in a couple of years
- Some general principles have been time-tested...

The commandments of the ninja

1. Thou should always document your work.
2. Thou should always use `plain text` documents.
3. Thy results should be reproducible.
4. Thy procedures should be reusable.
5. Thy code must be clean and clear.
6. Everything should be a script $\equiv$ the mouse is thy enemy.
7. Backup and move to the cloud.
8. Collaborate with colleagues and share code. Do no reinvent.
9. Thou shall choose open source applications (if possible).
10. Thy efforts of today shall make thou more productive in the long term.

2. The path of the ninja

White belt: the start of the journey

White belt: the start of the journey

Mindset

You have your whole scientific career in front of you and wonder if you could make decisions that would help you be more productive and, ultimately, become a better scientist.

White belt: the start of the journey

Mindset

You have your whole scientific career in front of you and wonder if you could make decisions that would help you be more productive and, ultimately, become a better scientist.

Tools

You love to use `MS Word` to create nice documents, `MS Excel` to analyse data and `MS Powerpoint` to create nice-looking figures.

You program in `Matlab`.

All your data is in your computer.

You have advanced applications with lots and lots of options that can do the most amazing tasks and you know where to find these hidden commands within the application menus.

White belt: the start of the journey

Important decision

Do you want to use your computer as a layperson or do you want to start the way of the ninja?

Yellow belt: scratching the surface

Yellow belt: scratching the surface

Mindset

You are decided to walk the path of the ninja, but you do not know where to start from.

Yellow belt: scratching the surface

Mindset

You are decided to walk the path of the ninja, but you do not know where to start from.

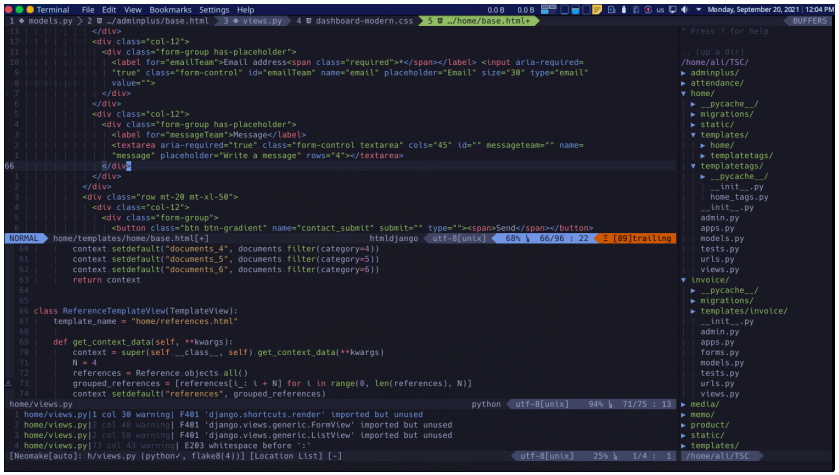
Tools

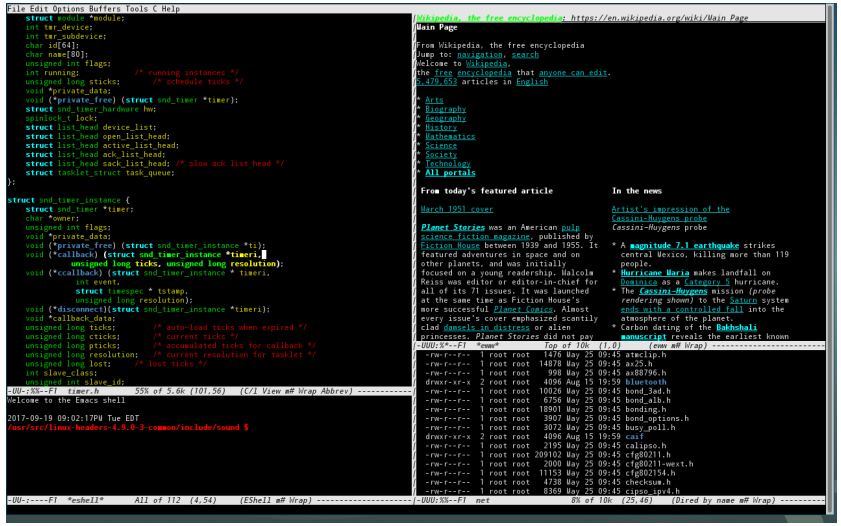
- You need an OS for scientific computing. Arguably, it is easier with a Linux or Mac Os machine.
- Learn urgently to work with the terminal, navigate the filesystem, see files, activate permissions, create links.
- You need to work remotely, either with `ssh` or an application. You also need to transfer files between your computer and remote ones; for that you need to learn `sftp` or an application that does it for you.

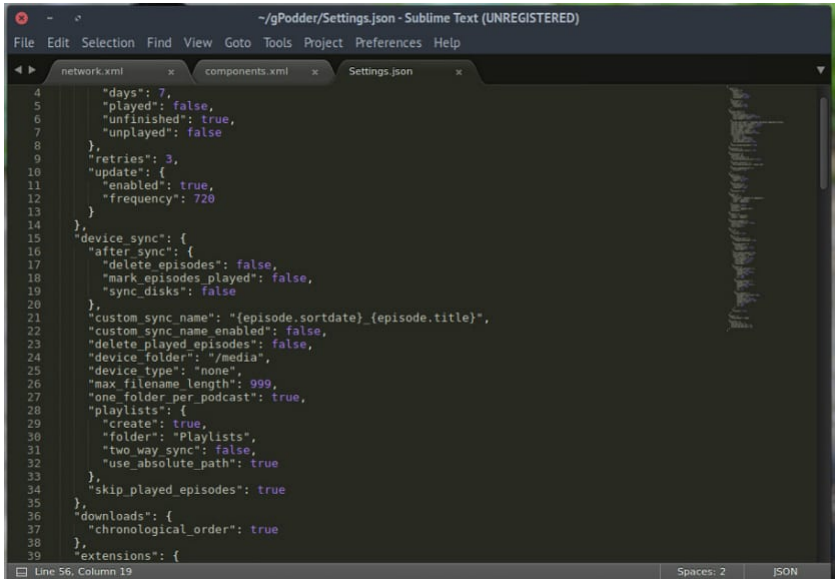
Yellow belt: scratching the surface

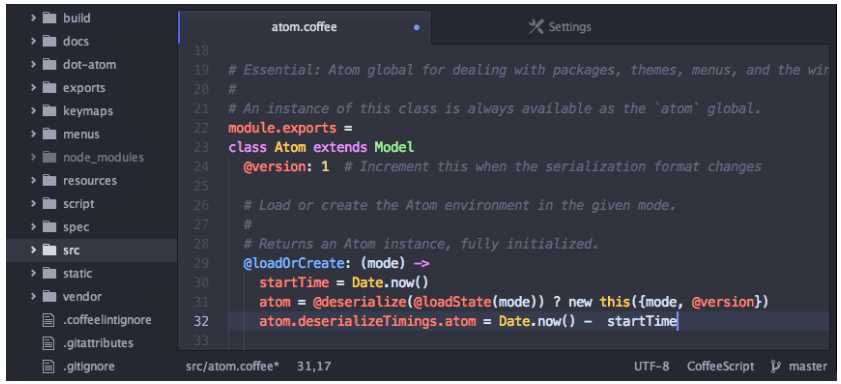
- Get a free account in `Dropbox`, `iCloud`, `oneDrive`, `ownCloud`, `Google drive` or any other cloud server to share documents.
- `Google docs` will allow you to work collaboratively on documents (proposals, papers, ...)
- Learn to use a text editor (`Word` is not a text editor!!):
  - unix-type: `Vi` , `Emacs`.
  - mac: `Sublime text`, `Atom` , ...
  - that can work well for all type of text files, color syntax, powerful add-ons.
  - It is a good thing if it can work both in unix and in mac/windows, so it can be used with the same configuration all types of machines.

Examples: editors









Yellow belt: scratching the surface

Important decision

Espouse an editor.
It is your most important application and the best friend of the ninja.
Spend time learning to become a power user.

Orange belt: waking up the beast

Orange belt: waking up the ninja in you

Mindset

- You can now write nice documents and do some computations, but you start to suspect that there might be better ways.
- You have seen around beautiful scientific documents.
- Up to now you had only written small code snippets, but now you are working on a complex code, and it's hard.
- Your advisor wants you to collaborate with her in writing a paper.

Orange belt: waking up the ninja in you

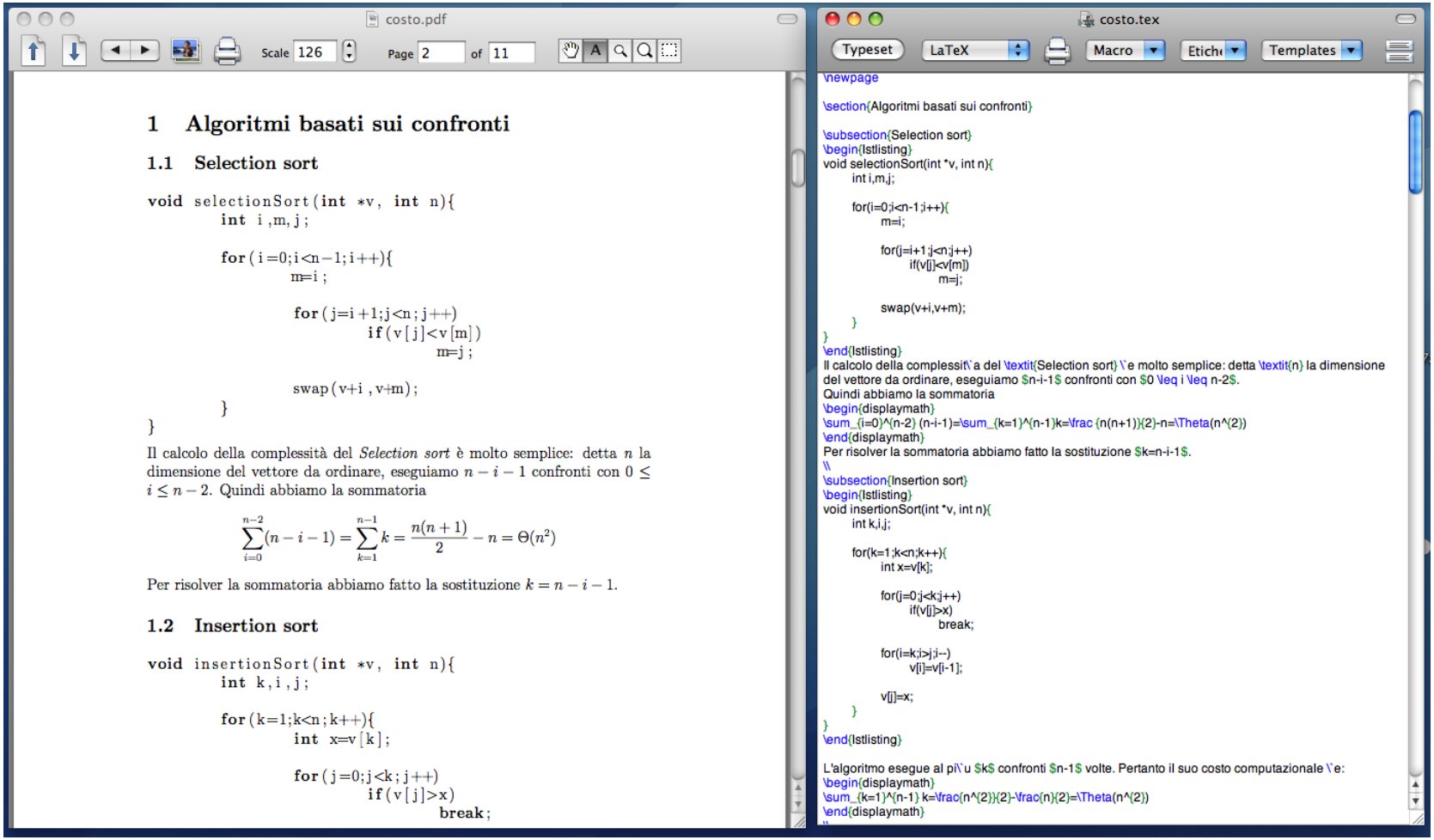
Tools for scientific writing

- LaTeX: get a distribution (`texmf` in linux/mac) or use `LyX`.
- `Overleaf` helps to collaborate in article preparation.

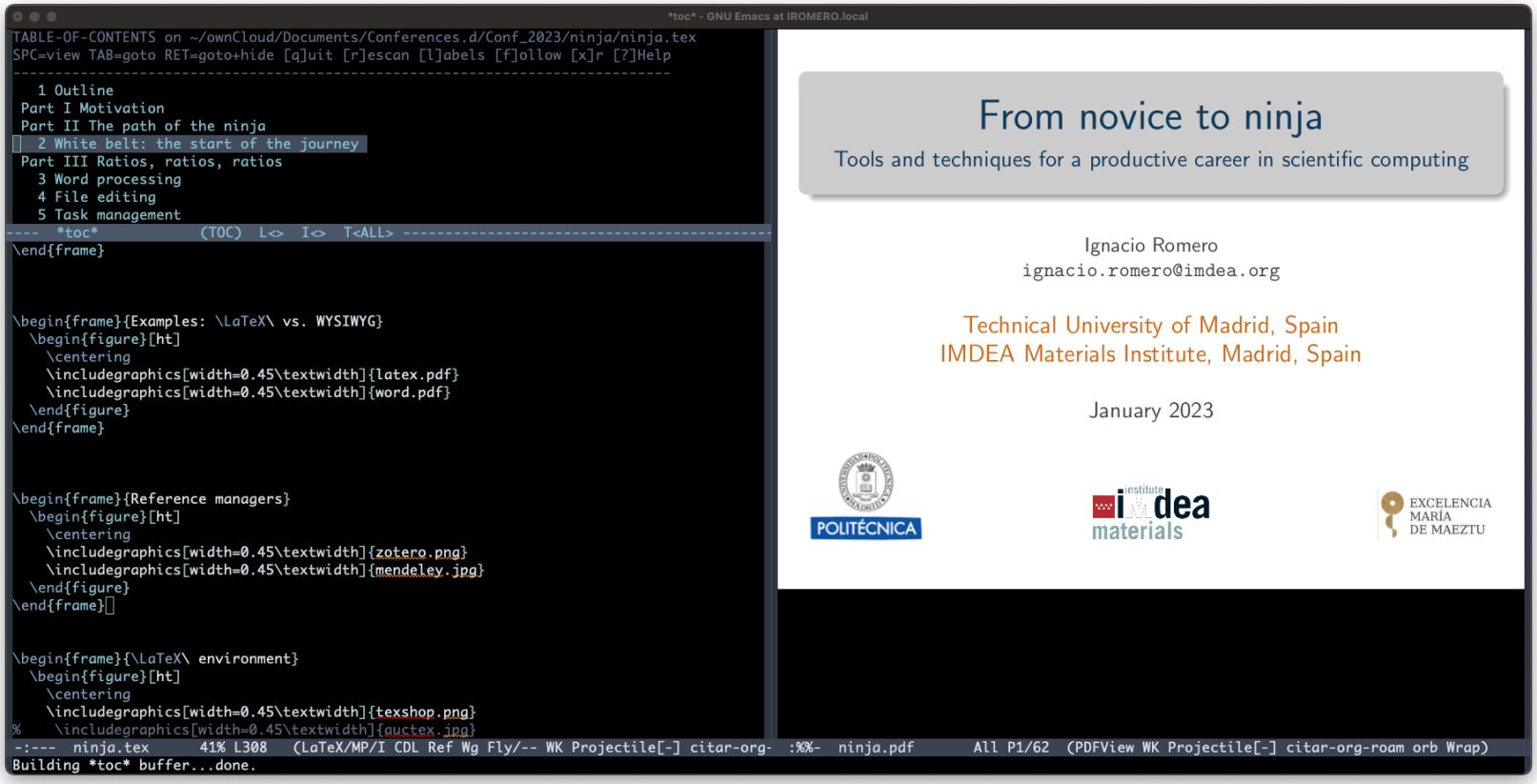
Important decision

Select a powerful environment for writing documents in LaTeX.

LaTeX environment



LaTeX environment



Examples: LaTeX vs. WYSIWYG

Notes on coupled stress-diffusion

I. Romero[1,2]

[1]Universidad Politécnica de Madrid, José Gutiérrez Abascal, 2, 28006 Madrid, Spain

[2] IMDEA Materials Institute, Eric Kandel,2; 28906 Madrid, Spain

June 3, 2022

1 Introduction

We summarize the equations that govern mass diffusion. Let χ be the concentration of a solute in a continuum. Then, Fick's law states that the mass flux $\boldsymbol{j}$ satisfies

$$\boldsymbol{j} = -A\nabla\chi \,,$$

for some diffusivity A. This is justified since the standard kinetic relation is

$$\boldsymbol{j} = -B\chi\nabla\mu \,,$$

where μ is the chemical concentration. In equilibrium, $\mu = C\log\chi$ and Fick's law follows.

2 Physical quantities

These are some of the quantities that appear in the equations:

Name	Symbol	Dimension
Concentration	χ	mol/L^3
Chemical potential	μ	E/mol
Grand-canonical potential	G	E
Mass diffusivity	D	L^2/T
Mass expansion		1/(mol/L^3)
Mass flux	$\boldsymbol{j}$	mol/(L^2 T)
Mass mobility		mol^2/(F L^2 T)
Number of atoms/ mol		1

Table 1: Summary of physical quantities involved in the model.

3 Equilibrium thermodynamics

Let us assume that there is a *free energy* function $\psi = \hat{\psi}(\chi)$, convex on its argument. The Legendre transform G is the grand canonical potential defined by

$$G(\mu) = \max_{\chi}(\psi(\chi) - \chi\mu).$$

The following relations hold

$$\mu = \frac{\partial\psi}{\partial\chi} \,, \qquad \chi = -\frac{\partial G}{\partial\mu}.$$

One example of a possible free energy is

$$\phi(\chi) = \mu_0\chi + \alpha\chi\left(\log\frac{\omega\chi}{1+\omega\chi} + \frac{m}{\omega\chi}\right)$$

Here μ_0 is a reference chemical potential, ω is the volume of a mole of solvent, and m is the Flory-Huggins parameter.

Notes on coupled stress-diffusion

Ignacio Romero[1,2]
[1]Universidad Politécnica de Madrid, José Gutiérrez Abascal, 28006 Madrid
[2]IMDEA Materials Institute, Eric Kandel 2; 28906 Getafe, Madrid

June 3, 2022

1. Introduction

We summarize the equations that govern mass diffusion. Let χ be the concentration of a solute in a continuum. Then, Fick's law states that the mass flux $\boldsymbol{j}$ satisfies

$$j = -A\nabla\chi$$

for some diffusivity A. This is justified since the standard kinetic relation is

$$j = -B\,\chi\nabla\mu,$$

where μ is the chemical concentration. In equilibrium $\mu = C\,log\chi$ and Fick's law follows.

2. Physical quantities

These are some of the quantities that appear in the equations:

Name	Symbol	Dimension
Concentration	χ	mol/L^3
Chemical potential	μ	E/mol
Grand-canonical potential	G	E
Mass diffusivity	D	L^2/T
Mass expansion		1/(mol/L^3)
Mass flux	j	mol/(L^2 T)
Mass mobility		mol^2/(F L^2 T)
Number of atoms/ mol		1

3. Equilibrium thermodynamics

Let us assume that there is a *free energy* function $\psi(\chi)$, convex on its argument. The Legendre transform G is the grand canonical potential defined by

$$G(\mu) = \max_{\chi}(\psi(\chi) - \chi\mu)$$

Orange belt: waking up the ninja in you

Common issue

- Articles, reports, and other documents start to pile up in your drive.
- Should I classify them by year, author, topic, ...?

Orange belt: waking up the ninja in you

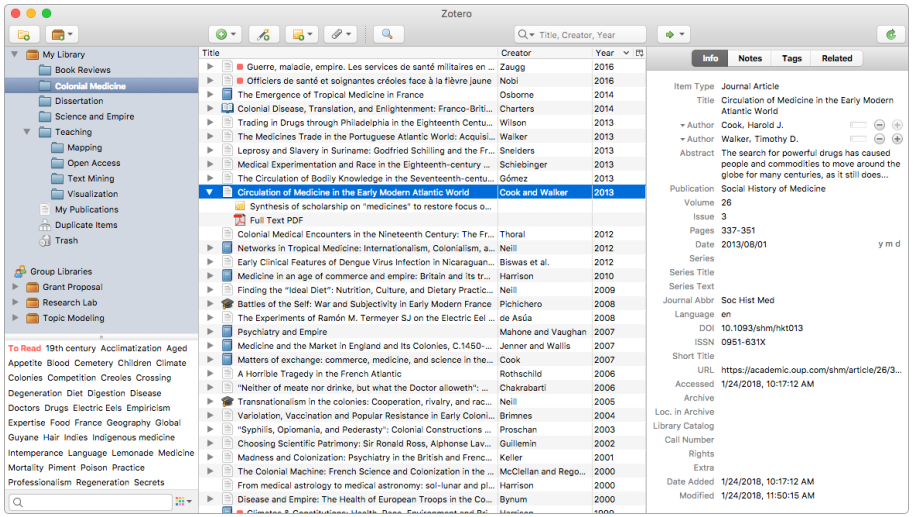
Common issue

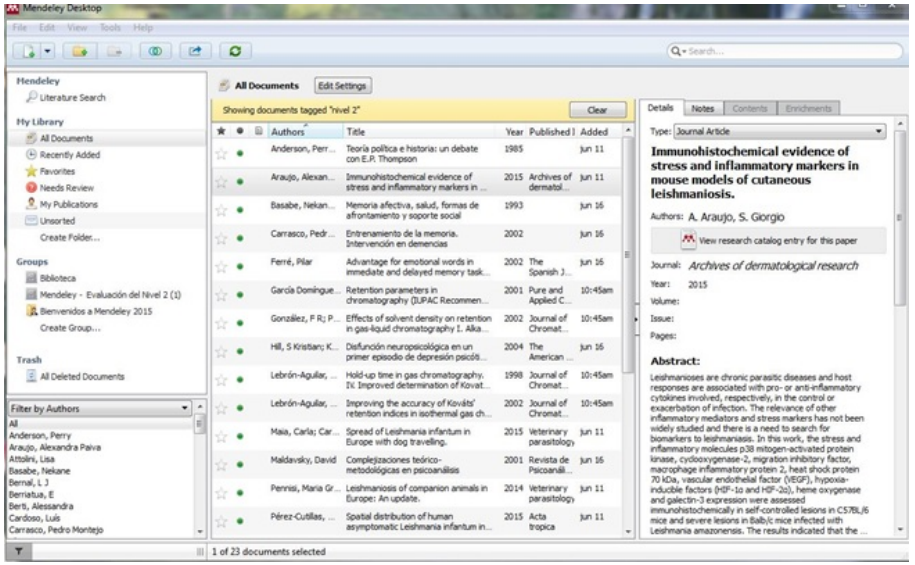
- Articles, reports, and other documents start to pile up in your drive.
- Should I classify them by year, author, topic, ...?

Tools for scientific writing

You create your own bibliographic DB and start using a reference manager (`Bibdesk`/`Mendeley`/`Zotero`...)

Reference managers





Orange belt: waking up the ninja in you

Common issue

- You have to redo the figures for a paper: you have received new data and have to start all over.
- You want to repeat the same figures you did for your experiments with steel, now for aluminum.

Orange belt: waking up the ninja in you

Common issue

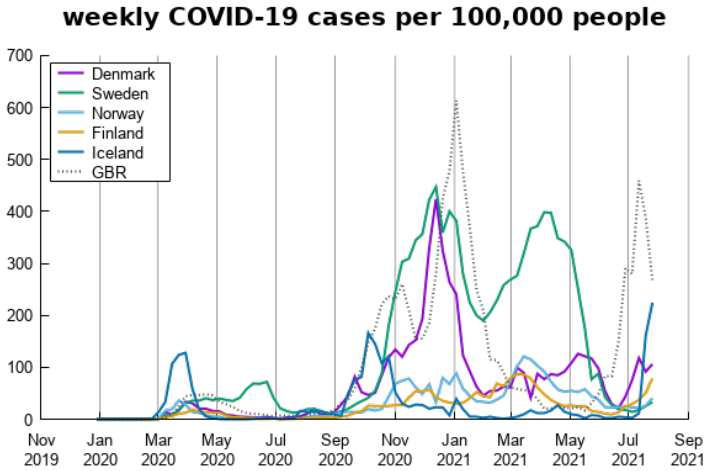
- You have to redo the figures for a paper: you have received new data and have to start all over.
- You want to repeat the same figures you did for your experiments with steel, now for aluminum.

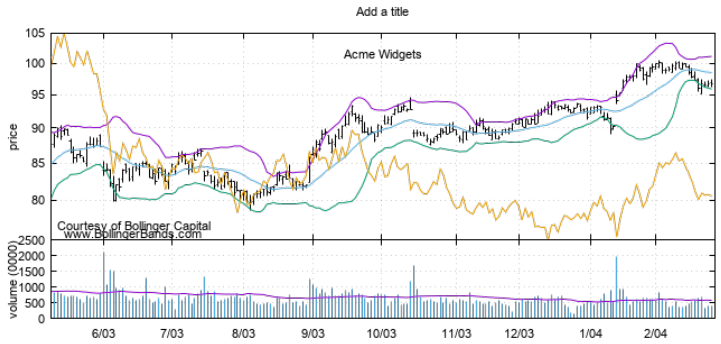
Tools for scientific writing

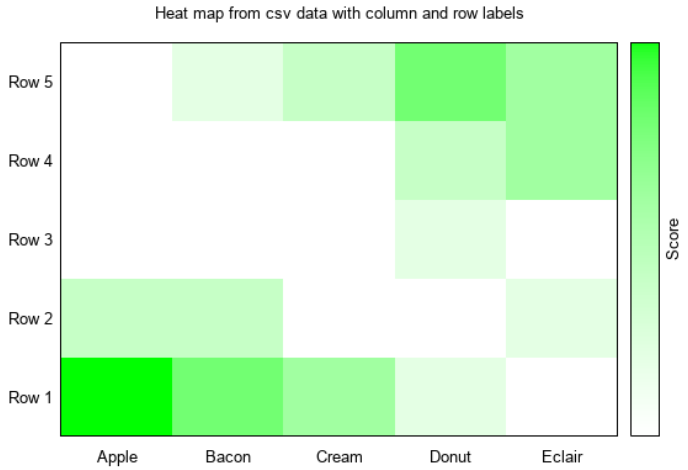
`Gnuplot` or `python` for scientific plotting: plots with scripts are re-usable and re-producible.

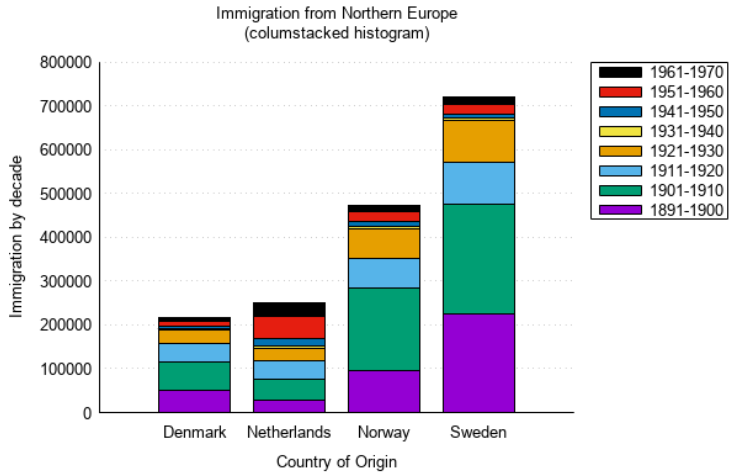
Examples: gnuplot

http://gnuplot.sourceforge.net/demo/









Orange belt: waking up the ninja in you

Tools for code development

- Learn to use an IDE (`Eclipse`, `Xcode`, `VisualStudio`, `Emacs`)
- Learn to debug!!

General tools

- Backup tools (brute force at least)
- Package management: linux: `apt-get`, macos: `darwinports` or `homebrew`
- Learn `emacs` or `vi`. You can go far with `nano` or `pico`, but not too far.

Orange belt: waking up the ninja in you

Tools for the big picture

- Statistical modelling: learn `R`

Green belt: standing on the shoulders of giants

Green belt: standing on the shoulders of giants

Mindset

- You have started to reap some of the benefits of advanced tools and are now decided to devote some serious time to learn not-so-simple things.
- You are starting to develop some taste for "nice" equations, figures, fonts.
- You wonder how is it possible that you ever used `MS Word`.

Green belt: standing on the shoulders of giants

Tools for code development

- `Python` programming and scripting.
- Version control with `Git` in `Github`, `Bitbucket`, `Gitlab`, ...

General tools

- `Makefiles`: automating tasks.
- `Shell` scripting.
- `cron`, scheduled tasks.

Green belt: standing on the shoulders of giants

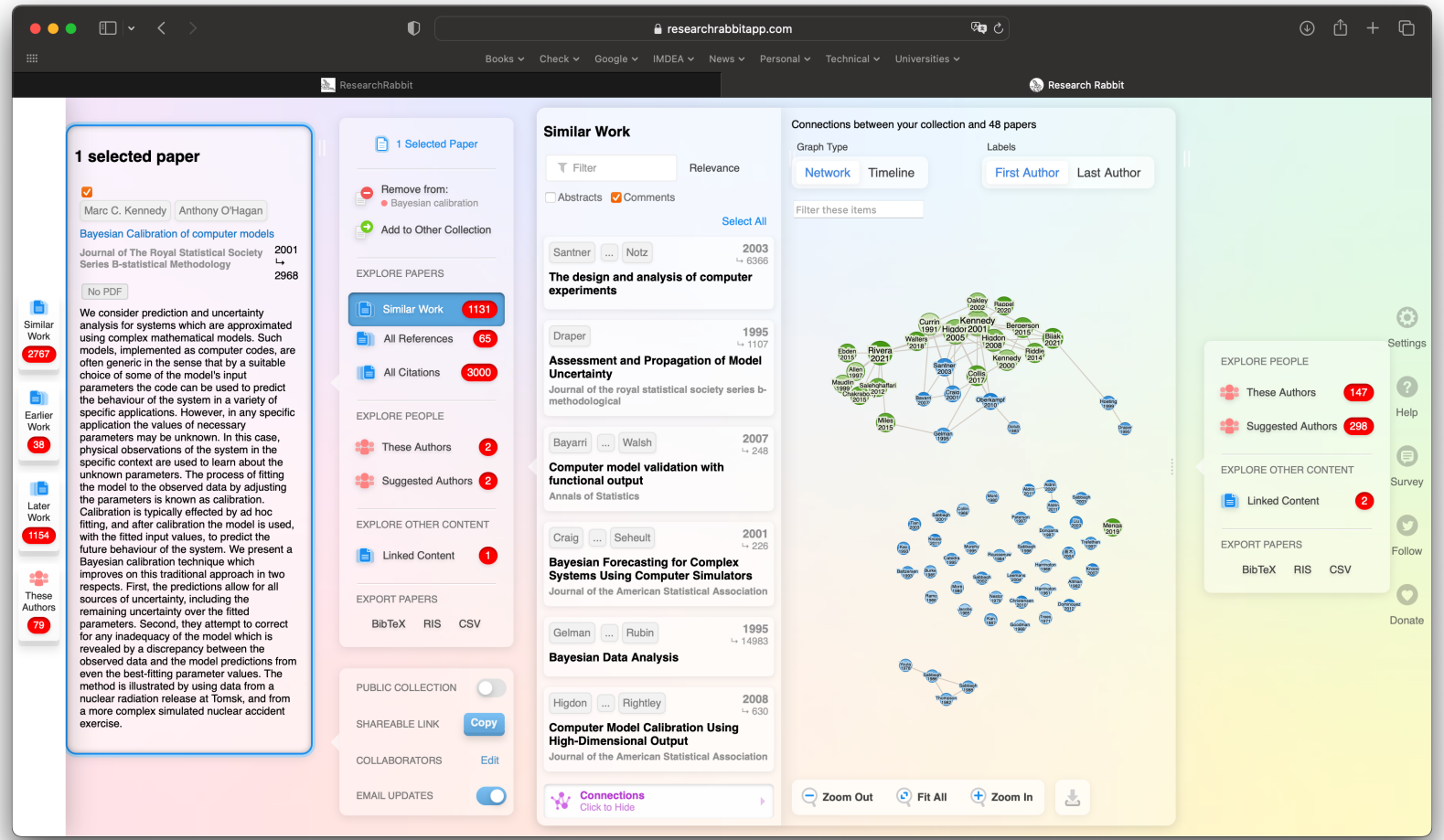
Tools for scientific writing

- `Beamer` for elegant presentation with lots of math formulas.
- Emacs `org-mode` for planning and literate programming.

Other scientific tools

- `Paraview`: a general purpose postprocessor
- `Gmsh`: a free mesh pre/post processor.
- `Numpy`, a `python` library for scientific computing
- `Research rabbit` or similar tool: structured bibliographic references.

Green belt: standing on the shoulders of giants



Blue belt: the force is in you

Blue belt: the force is in you

Mindset

- You are now half scientist, half hacker. You are amazed of the powerful things you can do with a computer, things that most mortals can't even dream of.

Blue belt: the force is in you

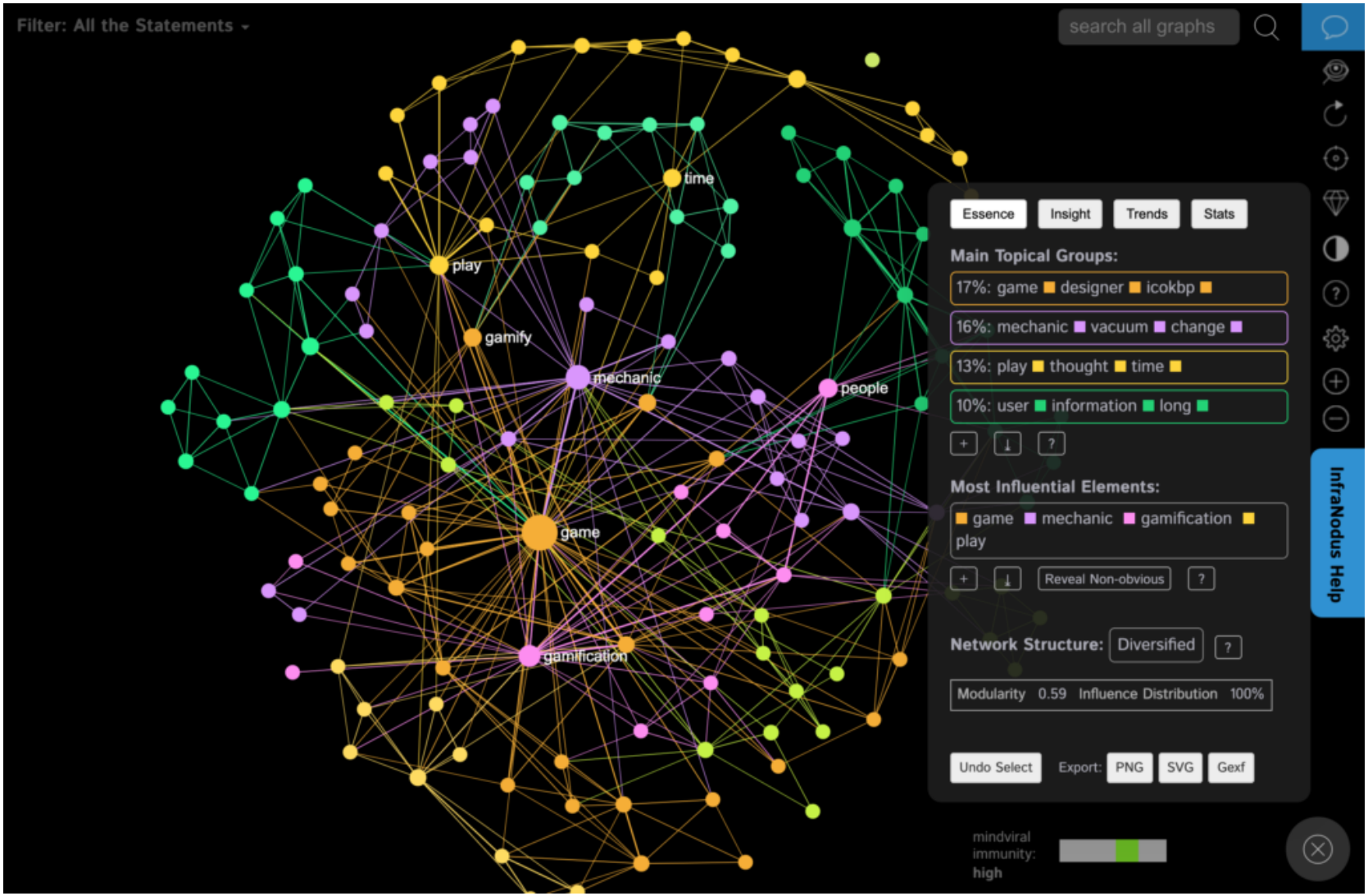
Mindset

- You are now half scientist, half hacker. You are amazed of the powerful things you can do with a computer, things that most mortals can't even dream of.

Tools

- Application profiling (timing, memory leaks, ...)
- Building and running codes in a MPI cluster.
- Knowledge base: `MS OneNote`, `Evernote`, `Roam`, `Notion`, `Obsidian`

Blue belt: the force is in you



Blue belt: the force is in you

Important decision

Make up your mind: commit or not commit to a knowledge based system. Another lifetime marriage.

Brown belt: bring it on!

Brown belt: bring it on!

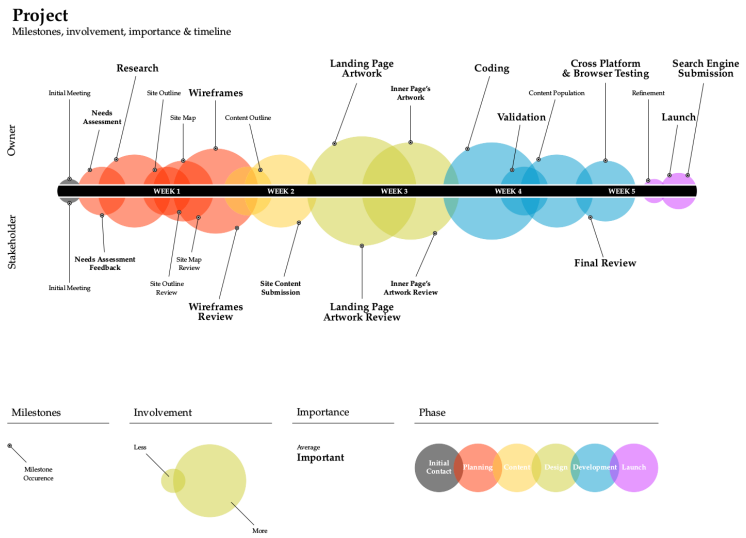
Mindset

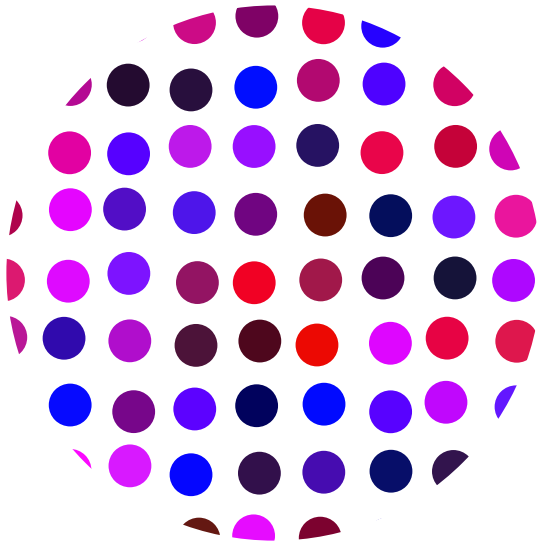
- You reached illumination: a computer is just an extension of the brain that can grow and make you faster, better organized, capable of storing and finding information. It helps you *think* better.

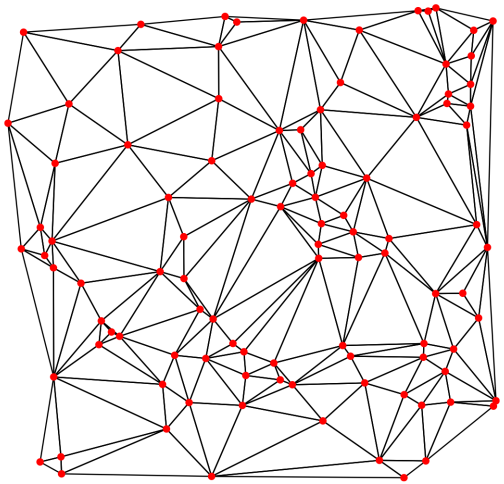
Tools

- You customize your `Beamer` package, for personalized presentations.
- Learn `Metapost` or `Asymptote`, programming languages for producing graphics.

Brown belt: bring it on!







Black belt: behold!

Black belt: behold!

Mindset

- You feel there is nothing that your computer can not help you to do better.
- You start to design new workflows, lifting most of the burdens from your shoulders, and delegating on your computer.

Tools

- You design your own `LaTeX` packages.
- You contribute to `Emacs` modes and packages.
- You create open-source tools for the community.

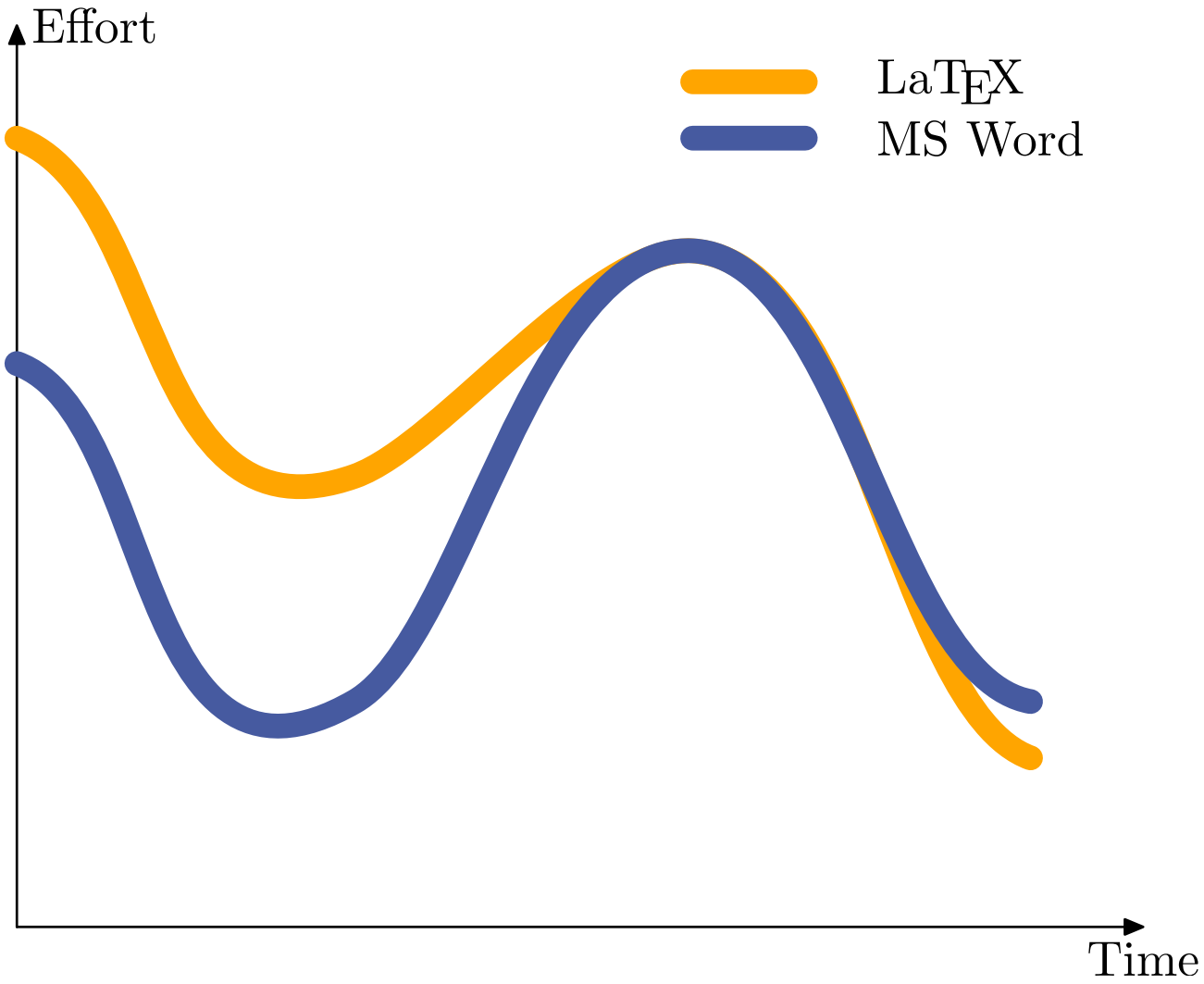
Black belt: behold!

Important decision

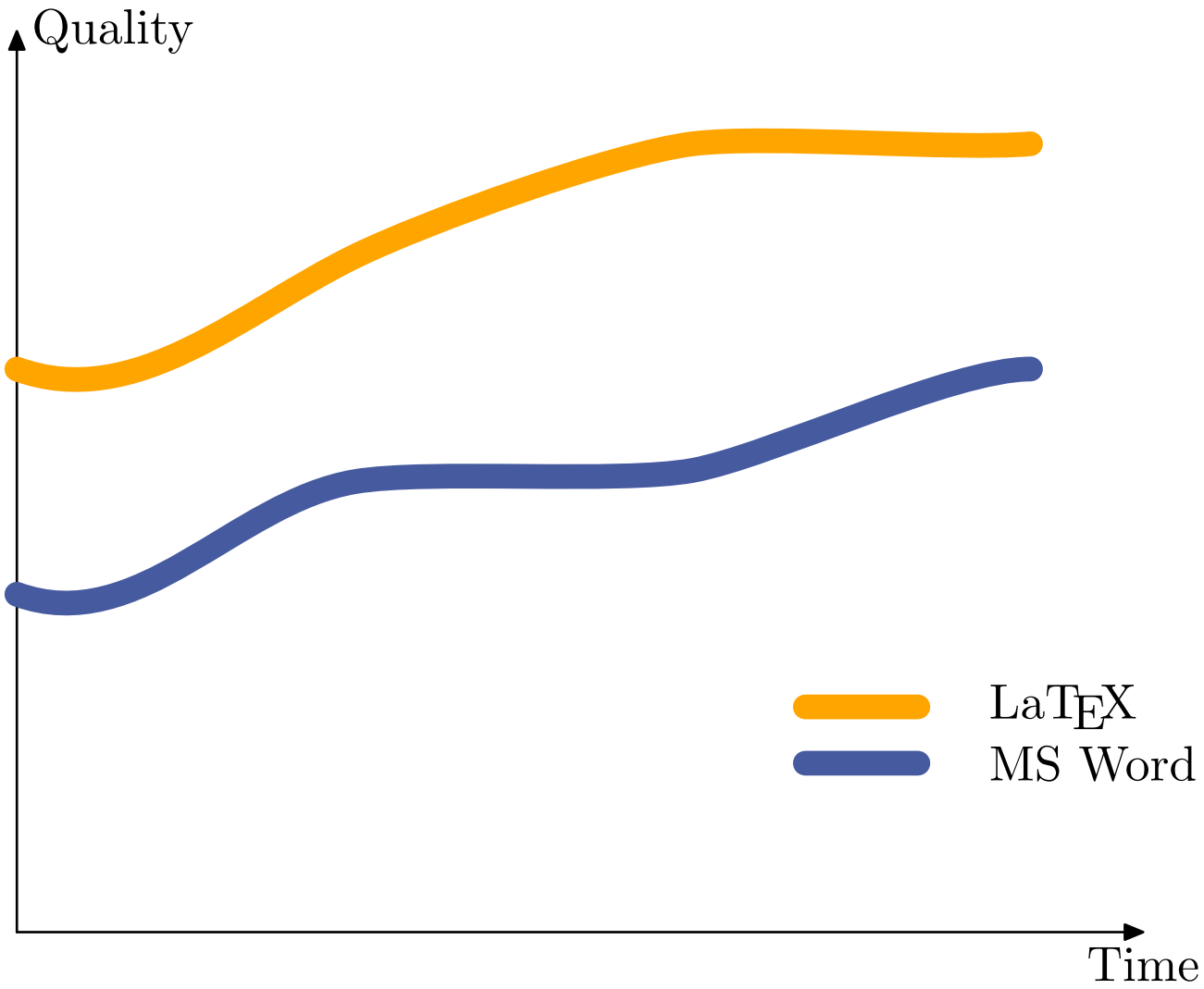
Since you have received so much, do you want to contribute and share?

3. Ratios, ratios, ratios

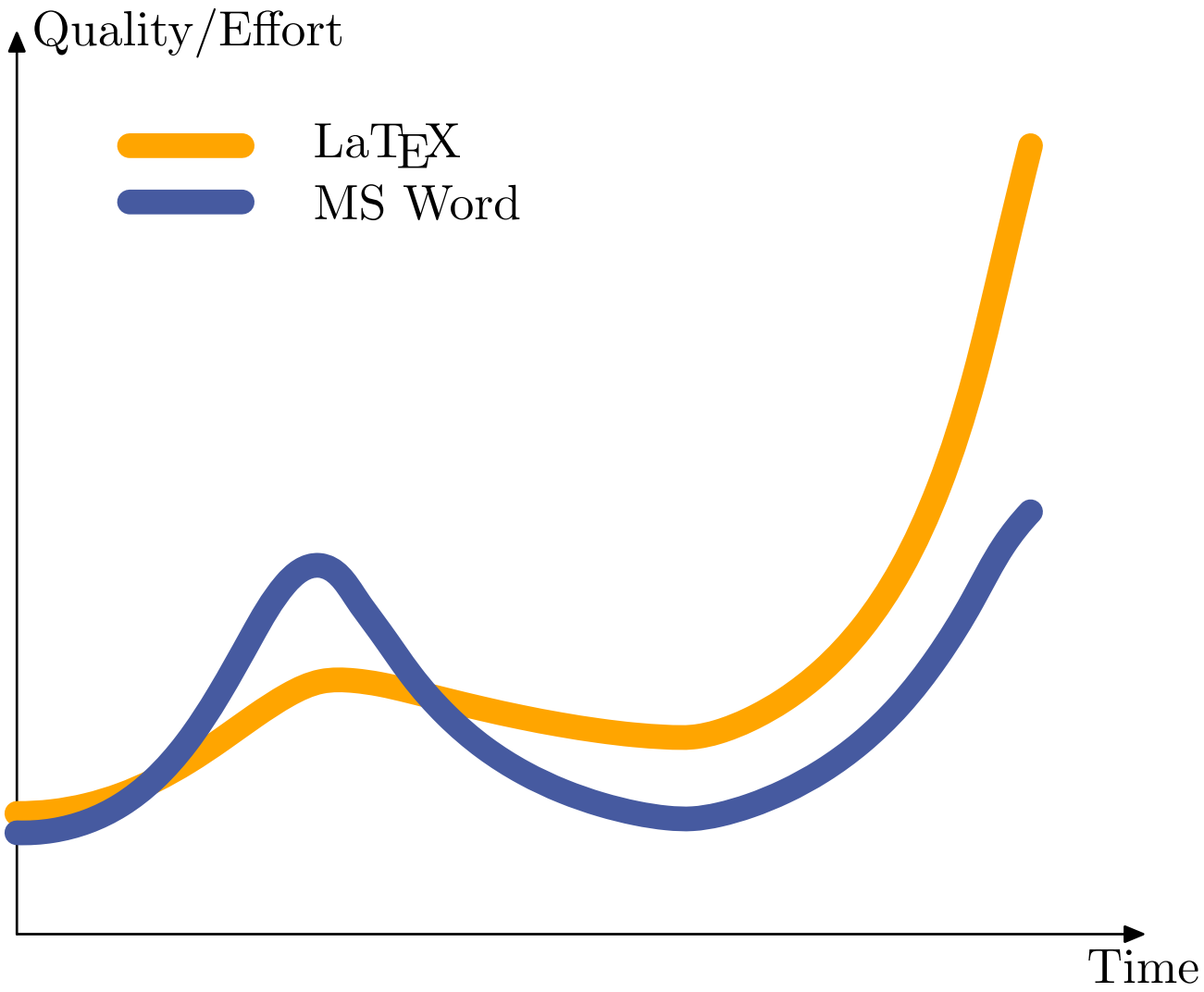
Word processing



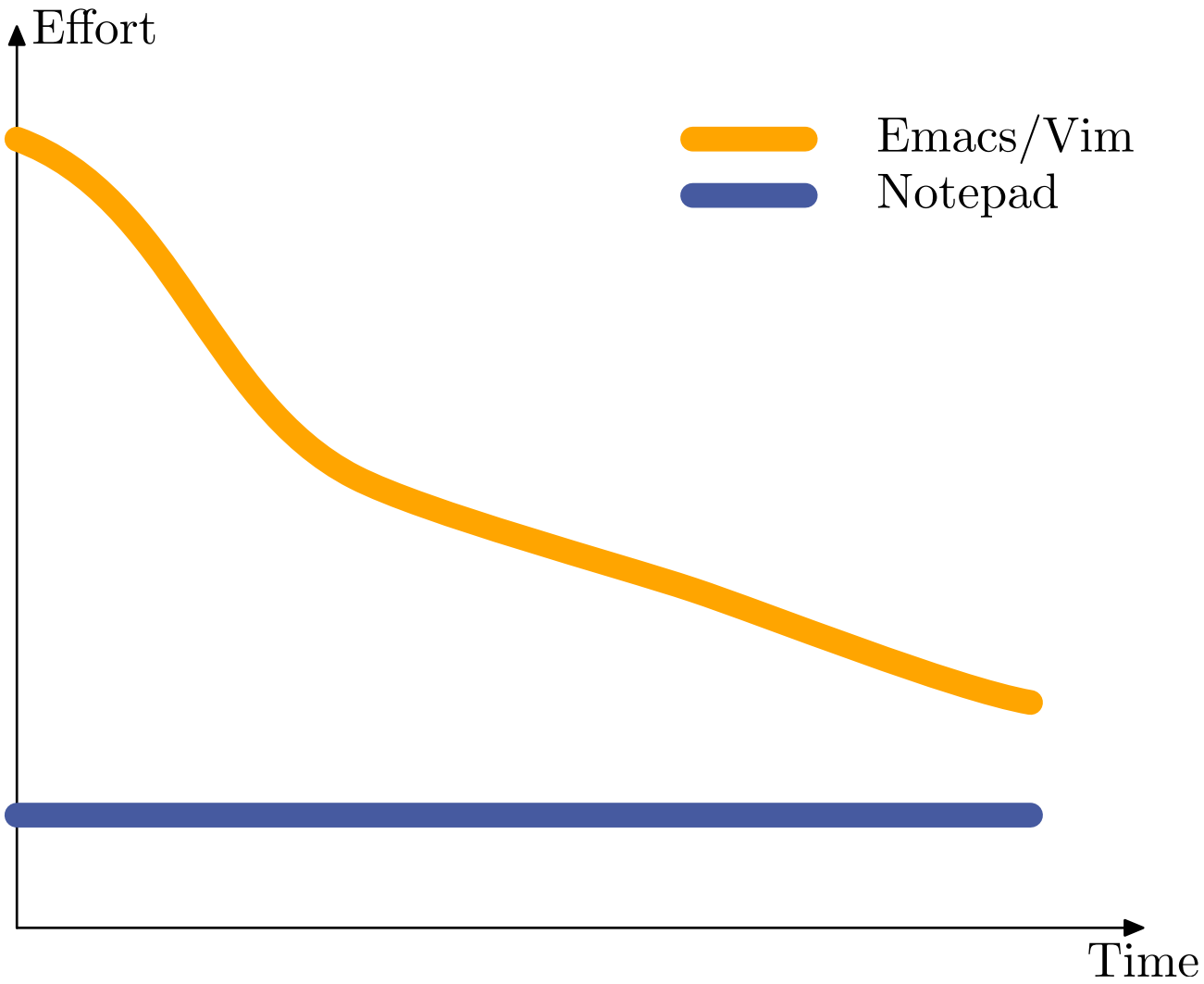
Word processing



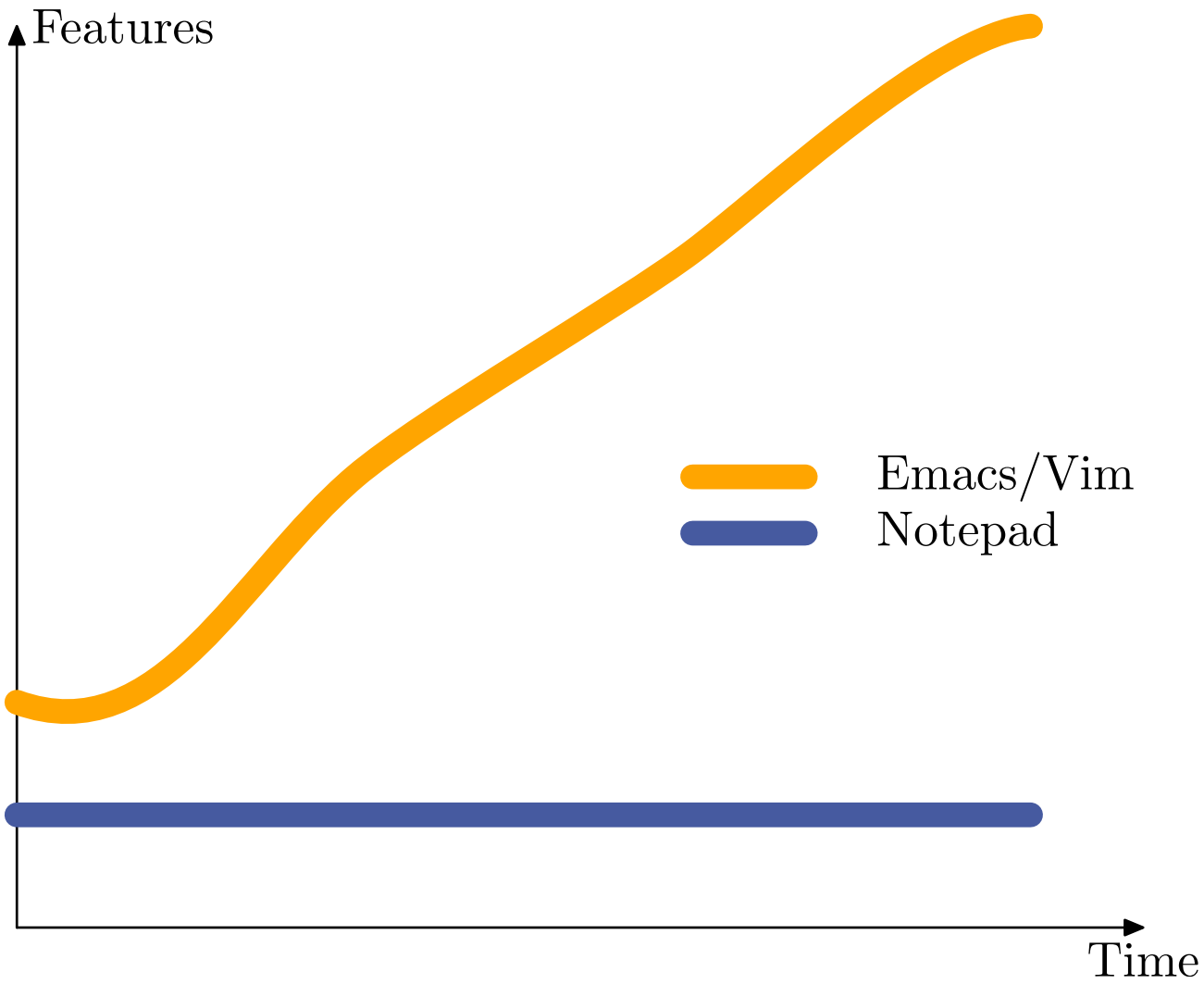
Word processing



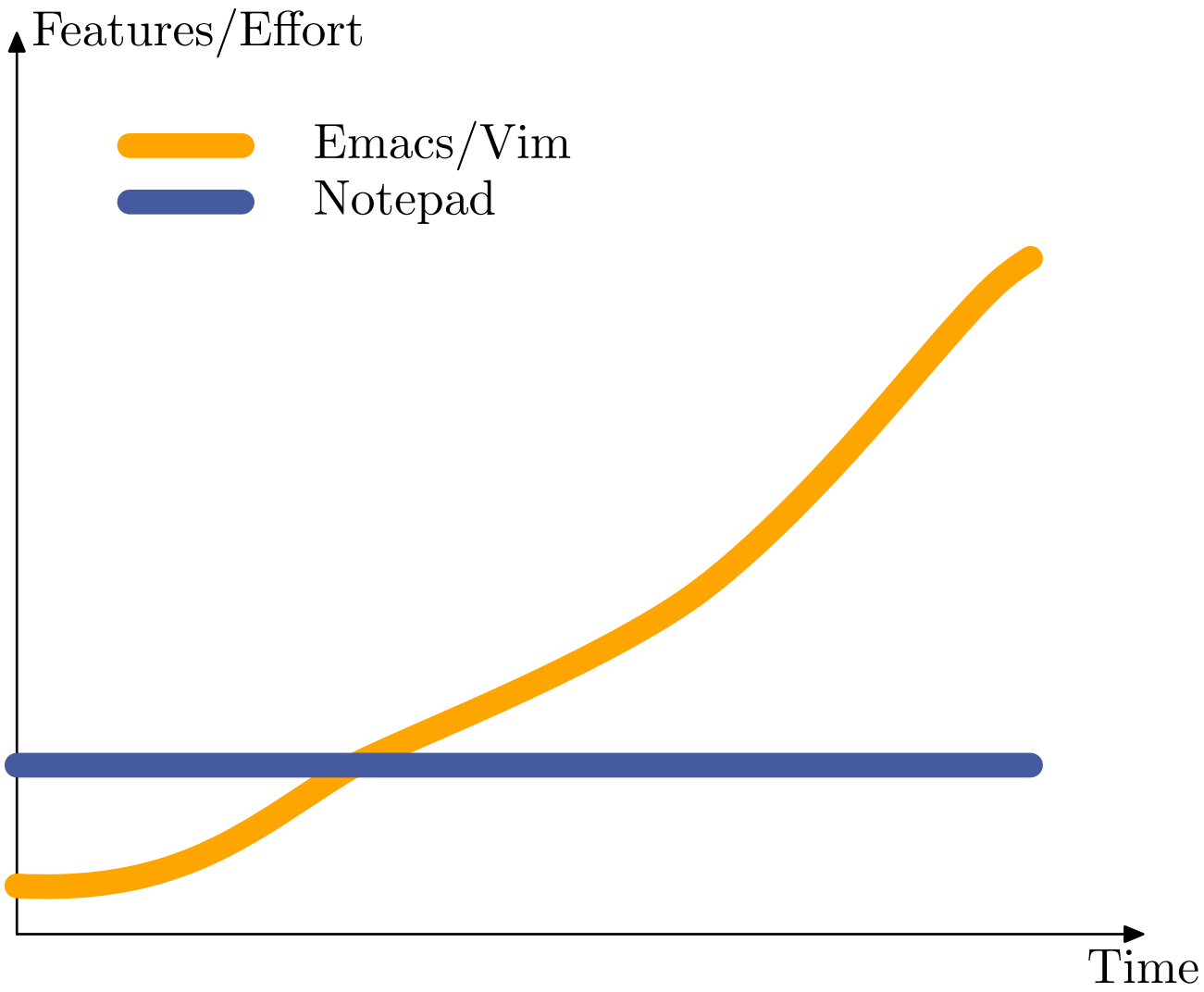
File editing



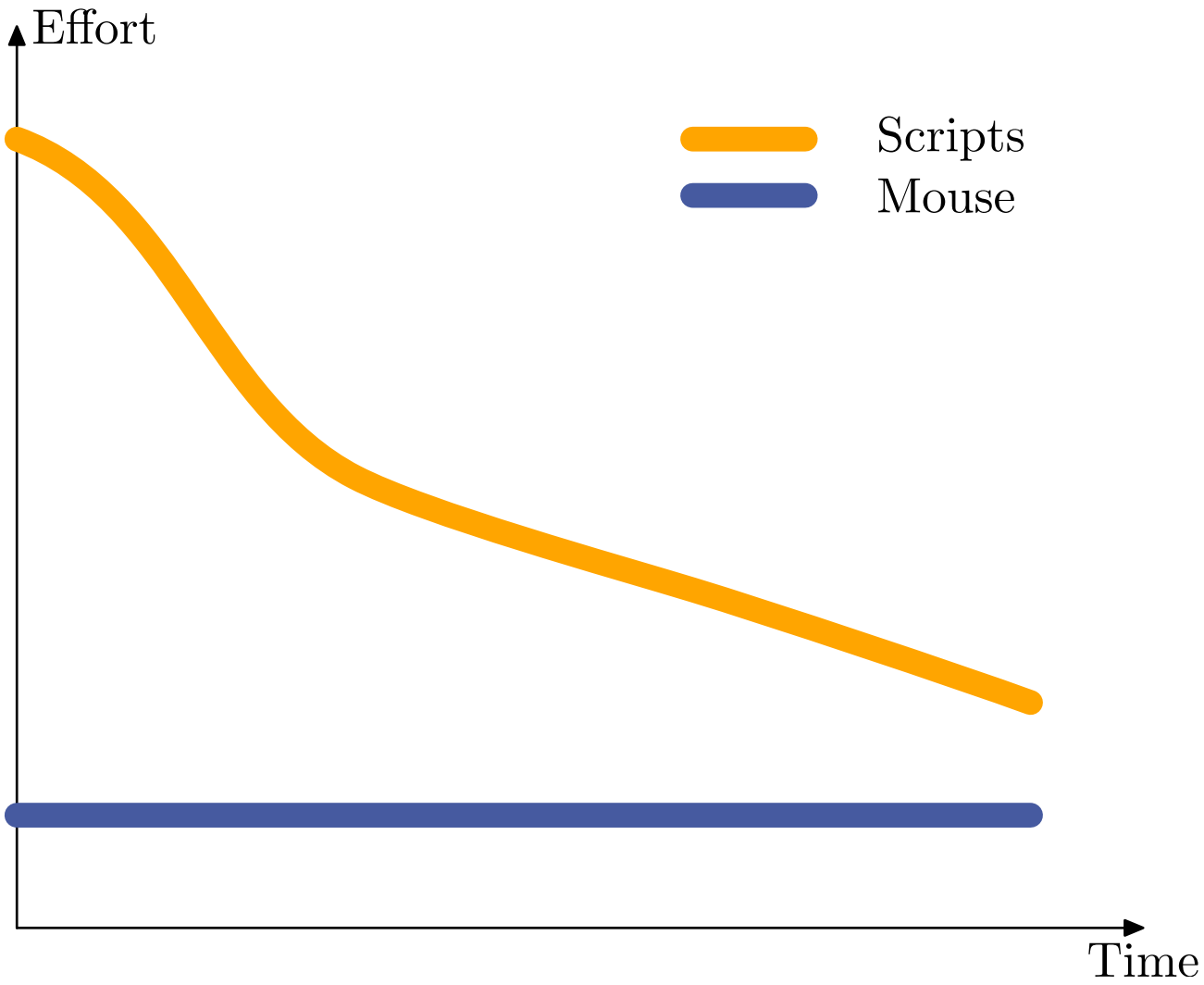
File editing



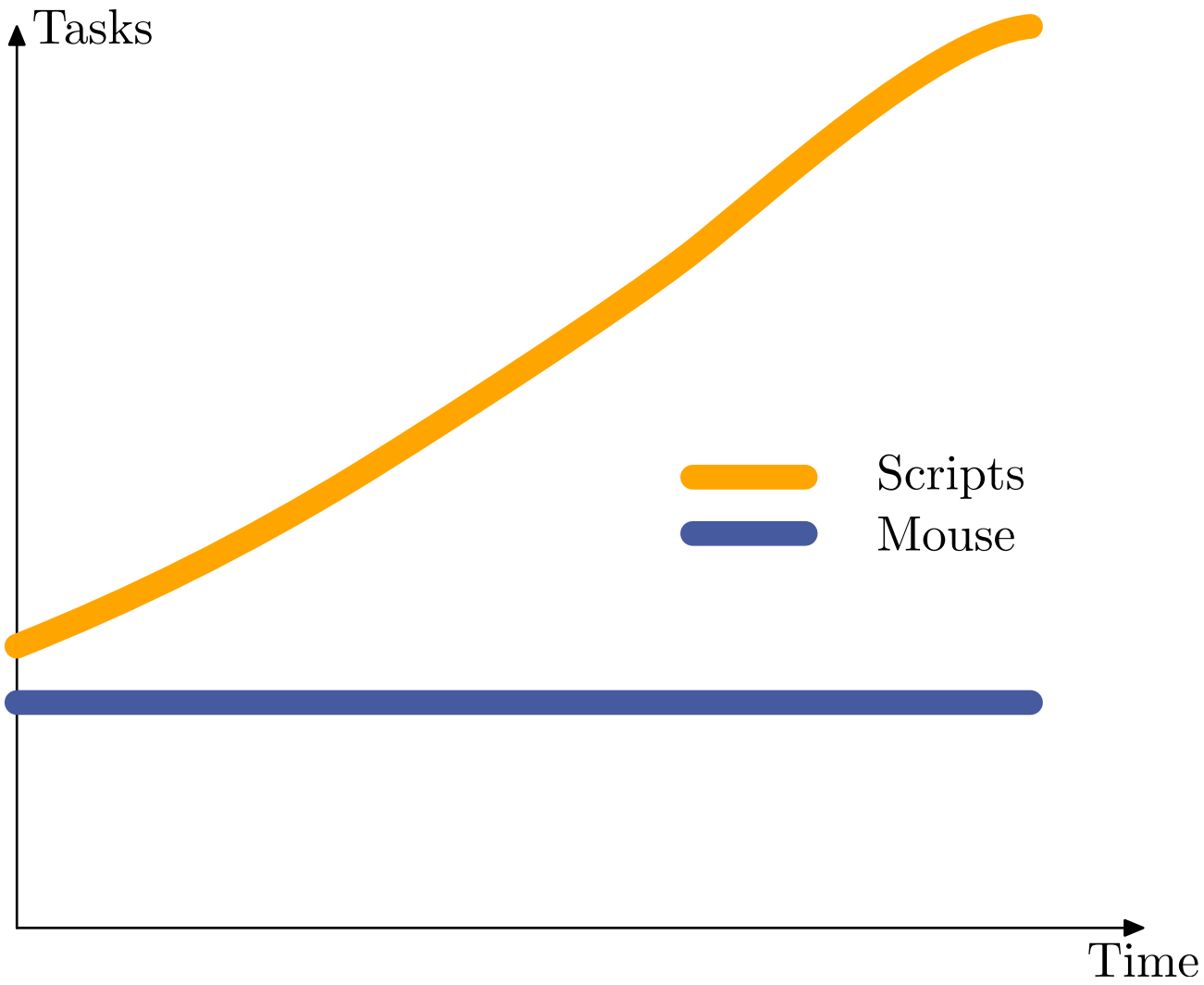
File editing



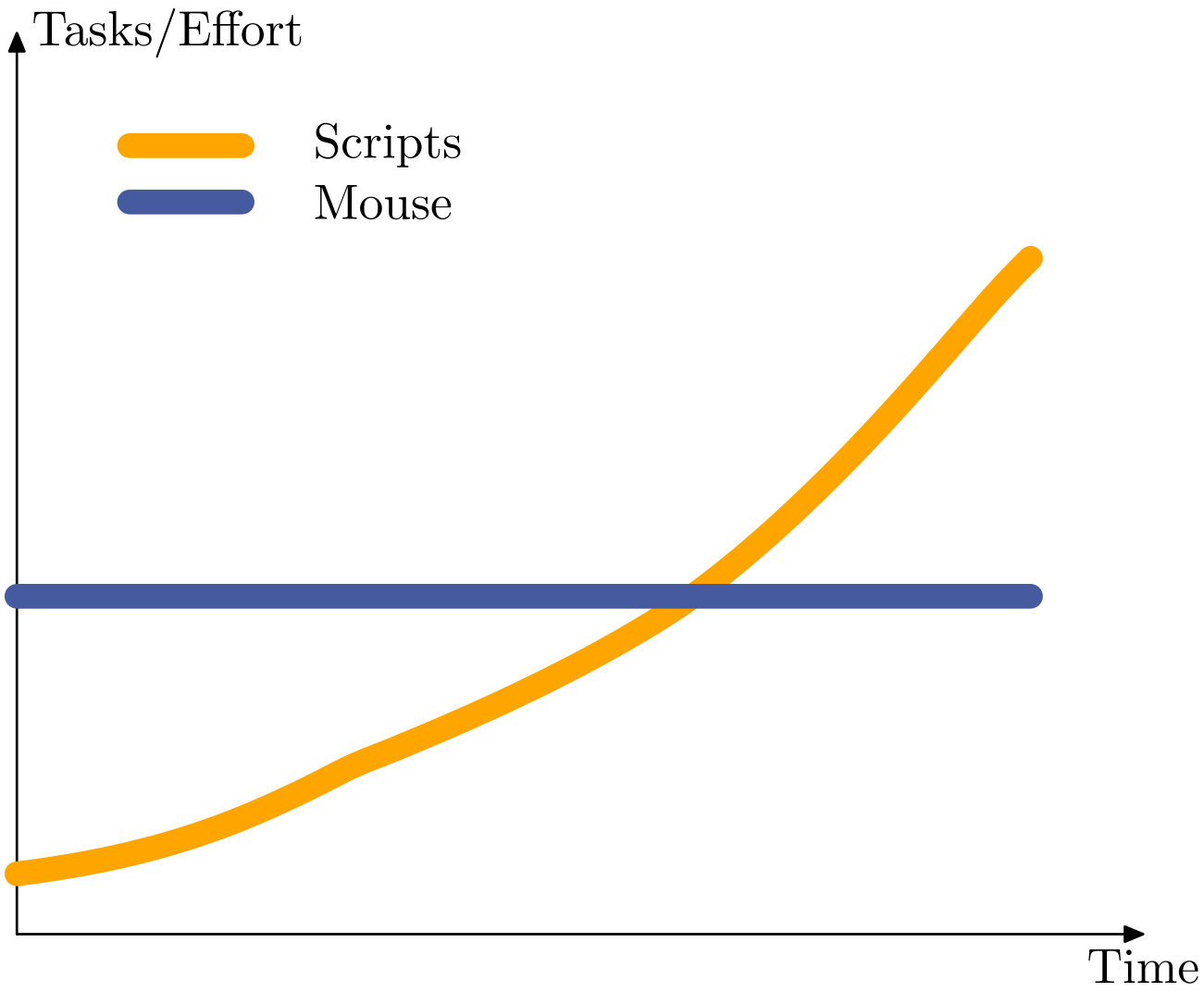
Task management



Task management



Task management



Some thoughts

- Powerful tools always demand more effort at the beginning.
- The right question is not *what is easier?* but rather *what option will give me the best outcome/effort ratio in the mid- to long-term?*
- There is not a single answer: it depends on the need for the numerator, your skill in the denominator, and your time frame.

4. Beyond

Final thoughts

- The computer should be like an instrument. The more you play it, the more beautiful things you can do with it, at a smaller cost.
- The computer should not be like a huge drawer, where you can dump and dump Gb of data and applications.
- There is much more: languages, contributing to the community, sharing code in public repositories, teaching, ...
- You are never done. This is a path ...

the path of the (scientific computing) ninja

5. External links

Operating system and tools

Basic Unix/Linux	https://www.linux.org
Sublime text	https://www.sublimetext.com
Atom	https://atom.io
Linux distro: Ubuntu	https://ubuntu.com
Homebrew	https://brew.sh
Vi	https://www.vim.org
Emacs	https://www.gnu.org/software/emacs
Git	https://git-scm.com
Advanced Emacs	https://www.gnu.org/software/emacs/

Scientific typesetting and graphics

LaTeX	https://www.vim.org
Lyx	https://www.lyx.org
Gnuplot	http://www.gnuplot.info
Beamer	http://tug.ctan.org/macros/latex/contrib/beamer
Metapost	https://tug.org/metapost.html
Asymptote	https://asymptote.sourceforge.io

Collaboration

Dropbox	https://www.dropbox.com/
Google docs	https://docs.google.com/
Overleaf	http://overleaf.com
`Github`	https://github.com
`Bitbucket`	https://bitbucket.com

Coding

Matlab	https://www.mathworks.com/products/matlab.html
Python	https://www.python.org
Eclipse	https://www.eclipse.org/ide/
Xcode	https://developer.apple.com/xcode/
Make(files)	https://www.gnu.org/software/make/manual/make.html
MPI	https://www.open-mpi.org
Emacs-lisp	https://www.gnu.org/software/emacs/

Bibliography and knowledge managers

Bibdesk	https://bibdesk.sourceforge.io
Zotero	https://www.zotero.org
Mendeley	https://www.mendeley.com
Emacs org-mode	https://orgmode.org
Roam	https://roamresearch.com
Notion	https://notion.so
Research Rabbit	https://www.researchrabbit.ai

Scientific computing and tools

Paraview	https://www.paraview.org
Gmsh	https://gmsh.info

From novice to ninja

Tools and techniques for a productive career in scientific computing

Ignacio Romero
ignacio.romero@imdea.org

Technical University of Madrid, Spain
IMDEA Materials Institute, Madrid, Spain

January 30th, 2023